

Lockheed  
Electronics  
Company, Inc.

A SUBSIDIARY OF  
LOCKHEED CORPORATION

1830 NASA Road 1, Houston, Texas 77058  
Tel. 713-333-5411

72  
JSC-14865

MAY 17 1979

CR160181

Ref: 642-7476  
Contract NAS 9-15800  
Job Order 73-705

TECHNICAL MEMORANDUM  
FISHER CLASSIFIER AND ITS PROBABILITY  
OF ERROR ESTIMATION

By

C. B. Chittineni

Approved By:

*T. C. Minter*  
T. C. Minter, Supervisor  
Techniques Development  
Section

(NASA-CR-160181) FISHER CLASSIFIER AND ITS  
PROBABILITY OF ERROR ESTIMATION (Lockheed  
Electronics Co.) 22 p HC A02/MF A01  
CSCL 12A N79-25813  
G3/64 23309 Unclass



May 1979

LEC-13301

## DISTRIBUTION

Distribution of this document is limited to those people whose names are followed by an asterisk in the following list; all others receive an abstract (JSC Form 1424) only.<sup>†</sup>

JSC/G. Badhwar/SF3*	NOAA/J. D. Hill/SF2
K. Baker/SF3*	D. G. McCrary/SF4
R. R. Baldwin/SF3	LEC/J. G. Baron
T. L. Barnett/SF3	M. L. Bertrand
R. M. Bizzell/SF4*	J. E. Davis
I. D. Browne/SF3	P. L. Krumm
L. F. Childs/SF2	D. G. Saile
K. J. Demel/SF3	P. C. Swanzy
H. G. deVezin/FM8	J. J. Vaccaro
J. W. Dietrich/SF3	Data Research and Control (3)*
J. L. Dragg/SF4*	Technical Library (5)*
R. B. Erb/SF1	Job Order File*
J. D. Erickson/SF3*	ERIM/Q. A. Holmes*
J. G. Garcia/SF3	R. Horvath
G. E. Graybeal/SF5	D. Rice
F. G. Hall/SF1*	KSU/A. M. Feyerherm
C. R. Hallum/SF4*	E. T. Kanemasu
K. E. Henderson/SF3	LARS/M. E. Bauer
W. E. Hensley/SF2	D. A. Landgrebe*
R. P. Heydorn/SF3*	T. L. Phillips
R. O. Hill/SF4	P. H. Swain
A. G. Houston/SF4	TAMU/L. F. Guseman
R. D. Juday/SF3	J. C. Harlan
T. W. Pendleton/SF3	H. O. Hartley
D. E. Pitts/SF3	UCB/R. N. Colwell
R. G. Stuff/SF3	C. M. Hay
D. R. Thompson/SF3	R. W. Thomas
M. C. Trichel/SF3*	UH/H. P. Decell
V. S. Whitehead/SF3	ESCS/W. H. Wigton
USDA/G. O. Boatwright/SF3	
A. D. Frank/SF221	
R. E. Hatch/SF221	
J. D. Murphy/SF221	
R. L. Packard/SF221	

---

<sup>†</sup>To obtain a copy of this document, contact one of the following:

J. D. Erickson — NASA/JSC Research, Test, and Evaluation Branch (SF3)

B. L. Carroll — LEC/SSD EO Development and Evaluation Department (626-42, C09)

1. Report No JSC-14865		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Fisher Classifier and Its Probability of Error Estimation				5. Report Date May 1979	
				6. Performing Organization Code	
7. Author(s) C. B. Chittineni				8. Performing Organization Report No LEC-13301	
9. Performing Organization Name and Address Lockheed Electronics Company, Inc. Systems and Services Division 1830 NASA Road 1 Houston, Texas 77058				10. Work Unit No. 63-1827-3705	
				11. Contract or Grant No NAS 9-15800	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Technical Monitor: J. D. Erickson/SF3				13. Type of Report and Period Covered Technical Memorandum	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract  This paper considers the Fisher classifier and the problem of estimating its probability of error. Computationally efficient expressions are derived for estimating the probability of error using the leave-one-out method. The optimal threshold for the classification of patterns projected onto Fisher's direction is derived. A simple generalization of the Fisher classifier to multiple classes is presented. Furthermore, computational expressions are developed for estimating the probability of error of the multiclass Fisher classifier.					
17. Key Words (Suggested by Author(s)) computationally efficient expressions, Fisher classifier, leave-one-out method, multiple classes, optimal threshold, probability of error				18. Distribution Statement	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 19	
22. Price*					

\*For sale by the National Technical Information Service, Springfield, Virginia 22161

## CONTENTS

Section	Page
1. INTRODUCTION . . . . .	1
2. FISHER CLASSIFIER . . . . .	2
3. RECURSIVE RELATIONS FOR THE FISHER WEIGHT VECTOR AND THRESHOLD . . . . .	3
4. SELECTION OF AN OPTIMAL THRESHOLD . . . . .	7
5. GENERALIZATION OF THE FISHER CLASSIFIER TO MULTIPLE CLASSES . . . . .	10
6. COMPUTATIONAL EXPRESSIONS FOR THE LEAVE-ONE-OUT METHOD IN A MULTICLASS CASE . . . . .	11
7. CONCLUSIONS . . . . .	14
8. REFERENCES . . . . .	15

## 1. INTRODUCTION

In the practical applications of pattern recognition, such as in remote sensing, there is considerable interest in the use of linear classifiers because they are simple and because fewer parameters need to be estimated. In many cases, it is required to estimate the probability of error in addition to designing the classifier. (For example in remote sensing, a separate set of labeled patterns is used in estimating the probability of error.) For designing the classifiers, the labels of the training patterns need to be obtained, and often acquiring labels is expensive. Hence, available training samples should be effectively used for designing the classifier and estimating the probability of error.

The leave-one-out method (ref. 1) is proposed in the literature as an effective way of estimating the probability of error from the training samples. The method is as follows. If there is a total of  $N$ -labeled patterns, leave out one pattern, design the classifier on remaining  $(N - 1)$  patterns, and test on the pattern that is left out. Repeat this process  $N$  times, every time leaving a different pattern, and then estimate the probability of error as an average of these errors. Use of this method, however, requires  $N$  classifiers to be designed. Fukunaga and Kessell (ref. 2) present a computational method for estimating the probability of error of a Bayes classifier using the leave-one-out method. Chittineni (ref. 3) developed a computational technique based on eigen perturbation theory for estimating the probability of error of the Fisher classifier using the leave-groups-out method.

This paper considers the Fisher classifier (refs. 4 and 5). The Fisher classifier is one of the most widely used linear classifiers. Computational expressions are developed based on matrix theory for estimating the probability of error of the Fisher classifier using the leave-one-out method. This paper is organized as follows.

Section 2 briefly presents the Fisher classifier. Section 3 develops computational expressions for using the leave-one-out method for estimating

Fisher's error probability. Section 4 discusses the effect of the Fisher threshold and presents expressions for obtaining the optimal threshold by minimizing the probability of error. Section 5 presents a simple generalization of the Fisher classifier to multiple classes. Section 6 develops computationally efficient expressions for the estimation of multicategory Fisher error using the leave-one-out method. Some matrix relations used in the paper are derived in the appendix (ref. 6).

## 2. FISHER CLASSIFIER

The Fisher classifier is a linear classifier that uses a direction  $W$  for the discriminant function,

$$g(X) = W^T X - t \quad (1)$$

so that when the training patterns are projected onto this direction, the intraclass patterns are clustered and the interclass patterns are separated to the extent possible as depicted in figure 1.

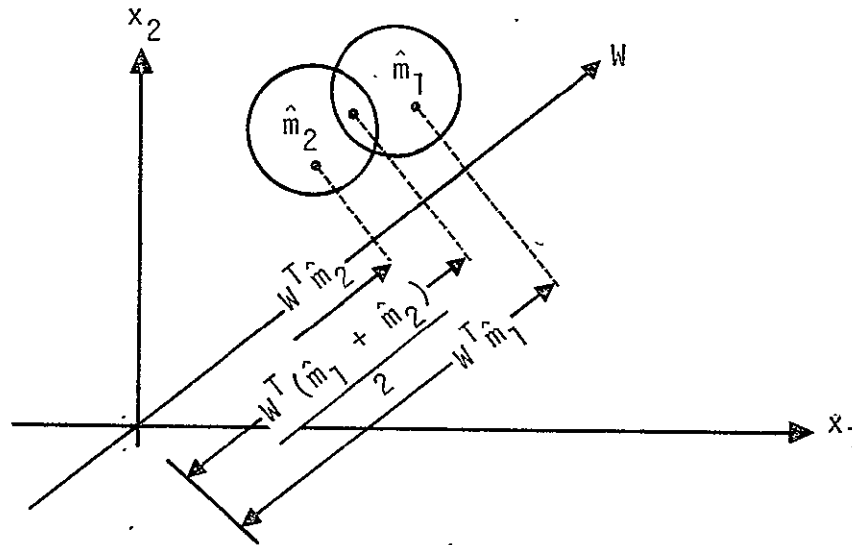


Figure 1.— Fisher's weight vector and threshold.

Let  $X_k^i \in \omega_i$ ,  $k = 1, 2, \dots, N_i$ ,  $i = 1, 2$  be the training pattern set. The unbiased estimates of means  $\hat{m}_i$  and covariance matrices  $\hat{\Sigma}_i$  of the patterns in the classes  $\omega_i$  are given by the following:

$$\hat{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_j^i \quad (2a)$$

$$\hat{\Sigma}_i = \frac{1}{(N_i - 1)} \sum_{j=1}^{N_i} (X_j^i - \hat{m}_i)(X_j^i - \hat{m}_i)^T \quad (2b)$$

The Fisher classifier chooses the weight vector  $W$ , such that the criterion  $\beta$  is maximized, where

$$\beta = \frac{[W^T(\hat{m}_1 - \hat{m}_2)]^2}{W^T \hat{S}_W W} \quad (3)$$

where  $\hat{S}_W = \hat{\Sigma}_1 + \hat{\Sigma}_2$ . The weight vector  $W$ , which maximizes  $\beta$ , can be shown to be

$$W = (\hat{S}_W)^{-1} (\hat{m}_1 - \hat{m}_2) \quad (4)$$

The Fisher threshold  $t$  is chosen as

$$t = W^T \frac{(\hat{m}_1 + \hat{m}_2)}{2} \quad (5)$$

The direction  $W$  and the threshold  $t$  are illustrated in figure 1. Fisher's decision rule is as follows:

$$\left. \begin{array}{l} \text{Decide } X \in \omega_1 \text{ if } g(X) > 0 \\ \text{Decide } X \in \omega_2 \text{ if } g(X) < 0 \end{array} \right\} \quad (6)$$

### 3. RECURSIVE RELATIONS FOR THE FISHER WEIGHT VECTOR AND THRESHOLD

In this section, computational expressions are developed for using the leave-one-out method with the Fisher classifier. The justification for the

leave-one-out method for estimating the probability of error is as follows. In general, the probability of error,  $\epsilon$ , is a function of two arguments:

$$\epsilon(\theta_1, \theta_2) \quad (7)$$

where  $\theta_1$  is the set of parameters for the distributions used to design the classifier and  $\theta_2$  is the set of parameters for the distributions used to test the performance. Let  $\theta$  and  $\hat{\theta}$  be the set of true parameters and their estimates. The  $\hat{\theta}$  is a random vector that depends on the particular sample used in its estimation. Let  $\hat{\theta}_N$  be a particular value of  $\hat{\theta}$ . Then (from ref. 7),

$$\epsilon(\theta, \theta) \leq \epsilon(\hat{\theta}_N, \theta) \quad (8)$$

Taking expectations on both sides, one gets

$$\epsilon(\theta, \theta) \leq E[\epsilon(\hat{\theta}_N, \theta)] \quad (9)$$

One of the ways of estimating the quantity on the RHS of equation (9) is with the leave-one-out method described in section 1. Presented in the following paragraphs are computational expressions for implementing the leave-one-out method with the Fisher classifier described in section 2. The cases in which a pattern  $x_k^1$  from class  $\omega_1$  is left out and in which a pattern from class  $\omega_2$  is left out can be treated similarly.

Let a pattern  $x_k^1$  from class  $\omega_1$  be left out and the patterns from class  $\omega_2$  remain. The means  $\hat{m}_i$ ,  $i = 1, 2$  and the covariance matrix  $\hat{\Sigma}_2$  are defined as in equations (2a) and (2b). Define the covariance matrix  $\hat{\Sigma}_1$  of the total pattern set from class  $\omega_1$  as

$$\hat{\Sigma}_1 = \frac{1}{(N_1 - 2)} \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_1)(x_j^1 - \hat{m}_1)^T \quad (10)$$

Let

$$\hat{S}_W = \hat{\Sigma}_1 + \hat{\Sigma}_2. \quad (11)$$



Note that  $\hat{\Sigma}_1$  is defined differently from the usual unbiased estimate for covariance matrices for mathematical simplicity; this definition will not affect the results. Now compute  $W$  and  $t$  as

$$W = \hat{S}_W^{-1}(\hat{m}_1 - \hat{m}_2) \quad (12)$$

and

$$t = W^T \frac{(\hat{m}_1 + \hat{m}_2)}{2} \quad (13)$$

When a pattern  $x_k^1$  from class  $\omega_1$  is left out, the unbiased estimates of the mean,  $\hat{m}_{1k}$  and the covariance matrix  $\hat{\Sigma}_{1k}$  of the patterns in class  $\omega_1$  are given by the following:

$$\hat{m}_{1k} = \frac{1}{(N_1 - 1)} \sum_{\substack{j=1 \\ j \neq k}}^{N_1} x_j^1 \quad (14)$$

and

$$\hat{\Sigma}_{1k} = \frac{1}{(N_1 - 2)} \sum_{\substack{j=1 \\ j \neq k}}^{N_1} (x_j^1 - \hat{m}_{1k})(x_j^1 - \hat{m}_{1k})^T \quad (15)$$

Let  $\hat{S}_{W1k} = \hat{\Sigma}_{1k} + \hat{\Sigma}_2$ . Then the Fisher weight vector  $W_{1k}$  and threshold  $t_{1k}$ , when a pattern  $x_k^1$  from class  $\omega_1$  is left out, are given by

$$W_{1k} = \hat{S}_{W1k}^{-1}(\hat{m}_{1k} - \hat{m}_2) \quad (16)$$

$$t_{1k} = \frac{W_{1k}^T(\hat{m}_{1k} + \hat{m}_2)}{2} \quad (17)$$

Expressions are now developed for the computation of  $W_{1k}$  and  $t_{1k}$  in terms of  $W$  and  $t$ . The relationships between  $\hat{m}_{1k}$ ,  $\hat{\Sigma}_{1k}$ ,  $\hat{S}_{W1k}$  and  $\hat{m}_1$ ,  $\hat{\Sigma}_1$ , and  $\hat{S}_W$  can be shown to be as follows (see the appendix):

$$\hat{m}_{1k} = \hat{m}_1 - \frac{1}{(N_1 - 1)}(x_k^1 - \hat{m}_1) \quad (18)$$

$$\hat{\Sigma}_{1k} = \hat{\Sigma}_1 - \frac{N_1}{(N_1 - 1)(N_1 - 2)} (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \quad (19)$$

$$\hat{S}_{W1k} = \hat{S}_W - \frac{N_1}{(N_1 - 1)(N_1 - 2)} (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \quad (20)$$

From equation (18), one obtains

$$(\hat{m}_{1k} - \hat{m}_2) = (\hat{m}_1 - \hat{m}_2) - \frac{1}{(N_1 - 1)} (x_k^1 - \hat{m}_1) \quad (21)$$

From equation (20), one obtains (appendix)

$$\hat{S}_{W1k}^{-1} = \hat{S}_W^{-1} + \frac{\alpha \hat{S}_W^{-1} (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \hat{S}_W^{-1}}{1 - \alpha (x_k^1 - \hat{m}_1)^T \hat{S}_W^{-1} (x_k^1 - \hat{m}_1)} \quad (22)$$

where

$$\alpha = \frac{N_1}{(N_1 - 1)(N_1 - 2)} \quad (23)$$

Let

$$\gamma(x_k^1) = \hat{S}_W^{-1} (x_k^1 - \hat{m}_1) \quad (24)$$

$$\beta(x_k^1) = (x_k^1 - \hat{m}_1)^T \hat{S}_W^{-1} (x_k^1 - \hat{m}_1) \quad (25)$$

$$\nu(x_k^1) = 1 - \alpha \beta(x_k^1) \quad (26)$$

$$\hat{m} = \frac{(\hat{m}_1 + \hat{m}_2)}{2} \quad (27)$$

$$z(x_k^1) = \gamma^T(x_k^1) (\hat{m}_1 - \hat{m}_2) \quad (28)$$

Using the definitions of equations (23) to (28), one obtains the following.

$$\begin{aligned}
W_{1k} &= \hat{S}_{W1k}^{-1}(\hat{m}_{1k} - \hat{m}_2) \\
&= \left[ \hat{S}_W^{-1} + \frac{\alpha Y(X_k^1) Y^T(X_k^1)}{v(X_k^1)} \right] \left[ (\hat{m}_1 - \hat{m}_2) - \frac{1}{(N_1 - 1)} (X_k^1 - \hat{m}_1) \right] \\
&= W - \frac{1}{(N_1 - 1) v(X_k^1)} Y(X_k^1) + \frac{\alpha Z(X_k^1)}{v(X_k^1)} Y(X_k^1)
\end{aligned} \tag{29}$$

$$\begin{aligned}
t_{1k} &= W_{1k}^T \frac{(\hat{m}_{1k} + \hat{m}_2)}{2} \\
&= t - \frac{W^T(X_k^1 - \hat{m}_1)}{2(N_1 - 1)} - \frac{Y^T(X_k^1) \hat{m}}{2(N_1 - 1) v(X_k^1)} + \frac{\beta(X_k^1)}{2(N_1 - 1)^2 v(X_k^1)} \\
&\quad + \frac{\alpha Z(X_k^1)}{v(X_k^1)} Y^T(X_k^1) \hat{m} - \frac{\alpha Z(X_k^1) \beta(X_k^1)}{2(N_1 - 1) v(X_k^1)}
\end{aligned} \tag{30}$$

Equations (29) and (30) can be used to compute  $W_{1k}$  and  $t_{1k}$  from  $W$  and  $t$ , every time that a pattern  $X_k^1$  is left out from class  $\omega_1$  and the pattern  $X_k^1$  is tested. Similarly, recursive expressions can be derived when a pattern  $X_k^2$  is left out from class  $\omega_2$ . It is to be noted that because the covariance matrices are defined as in equation (10), the matrix  $\hat{S}_W$  is to be computed and inverted twice, once when patterns from class  $\omega_1$  are left out and again when patterns from class  $\omega_2$  are left out.

#### 4. SELECTION OF AN OPTIMAL THRESHOLD

This section considers the problem of finding the optimum threshold,  $t$ , to achieve minimum probability of error for the projected patterns onto Fisher's direction. The patterns in class  $\omega_i$  are assumed to be normally distributed; i.e.,  $p(X|\omega_i) \sim N(m_i, \Sigma_i)$ . Let  $y$  be the projection of pattern  $X$  onto Fisher's direction  $W$ ; i.e.,

$$y = W^T X \tag{31}$$

Since  $X$  is normally distributed,  $y$  is also normally distributed; i.e.,

$$p(y|\omega_i) \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2 \quad (32)$$

where

$$\mu_i = W^T m_i \quad (33)$$

and

$$\sigma_i^2 = W^T \Sigma_i W \quad (34)$$

If Fisher's decision rule is used, decide  $y \in \omega_1$  if  $y > t$ ; otherwise decide  $y \in \omega_2$ , the probability of error incurred can be written as

$$\begin{aligned} P_e &= P_1 \int_{-\infty}^t p(y|\omega_1) dy + P_2 \int_t^{\infty} p(y|\omega_2) dy \\ &= P_1 \int_{-\infty}^{\frac{t-\mu_1}{\sigma_1}} \phi(\zeta) d\zeta + P_2 \int_{\frac{t-\mu_2}{\sigma_2}}^{\infty} \phi(\zeta) d\zeta \end{aligned} \quad (35)$$

where  $\phi(\zeta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \zeta^2\right)$  and  $P_i$  are the *a priori* probabilities of the classes  $\omega_i$ ,  $i = 1, 2$ . On differentiating equation (35) with respect to  $t$ , the following is obtained:

$$\frac{\partial P_e}{\partial t} = P_1 \phi\left(\frac{t - \mu_1}{\sigma_1}\right) \frac{1}{\sigma_1} - P_2 \phi\left(\frac{t - \mu_2}{\sigma_2}\right) \frac{1}{\sigma_2} \quad (36)$$

Equating  $\frac{\partial P_e}{\partial t}$  to zero and then simplifying it, one obtains

$$\left(\frac{t - \mu_2}{\sigma_2}\right)^2 - \left(\frac{t - \mu_1}{\sigma_1}\right)^2 = 2 \log\left(\frac{P_2 \sigma_1}{P_1 \sigma_2}\right) \quad (37)$$

The following cases are considered:

$$\text{Case (1): } P_1 = P_2, \sigma_1 = \sigma_2$$

Obtained from equation (37) is the optimum value of  $t$  that minimizes the probability of error for Fisher's direction as

$$t = \frac{\mu_1 + \mu_2}{2} \quad (38)$$

Equations (13) and (38) show that this is the threshold that is often implemented with the Fisher classifier.

Case (2):  $P_1 \neq P_2, \sigma_1 = \sigma_2 = \sigma$

In this case, the optimum value of threshold  $t$  can be obtained from equation (37) as

$$t = \frac{\sigma^2}{(\mu_1 - \mu_2)} \log\left(\frac{P_2}{P_1}\right) + \left(\frac{\mu_1 + \mu_2}{2}\right) \quad (39)$$

Case (3):  $P_1 \neq P_2, \sigma_1 \neq \sigma_2$

On simplification, the following is obtained from equation (37):

$$t^2 + \frac{(2\mu_1\sigma_2^2 - 2\mu_2\sigma_1^2)}{(\sigma_1^2 - \sigma_2^2)}t + \frac{(\sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2)}{\sigma_1^2 - \sigma_2^2} - \frac{2\sigma_1^2\sigma_2^2}{(\sigma_1^2 - \sigma_2^2)} \log\left(\frac{P_2}{P_1} \frac{\sigma_1}{\sigma_2}\right) = 0 \quad (40)$$

This is a quadratic equation of the form  $at^2 + bt + c = 0$ . The discriminant of the equation  $\eta = b^2 - 4ac$  can be shown to be

$$\eta = \frac{4}{\left(\frac{\sigma_1}{\sigma_2} - \frac{\sigma_2}{\sigma_1}\right)^2} \left[ (\mu_1 - \mu_2)^2 + 2(\sigma_1^2 - \sigma_2^2) \log\left(\frac{P_2}{P_1} \frac{\sigma_1}{\sigma_2}\right) \right] \quad (41)$$

From equation (41), it is seen that when  $P_1 = P_2$ ,  $\eta$  is always positive, thus giving real roots for equation (40). Even when  $P_1 \neq P_2$ , if  $\eta$  is positive, real roots are obtained for  $t$ . The  $\eta$  is negative when there exists no real threshold that minimizes the probability of error. Equation (40) gives two

roots for  $t$ . Since  $P_e$  is continuous in  $t$ , the  $t$  that minimizes  $P_e$  can be obtained by looking at the second derivative of  $P_e$ . Differentiating equation (36) with respect to  $t$ , one obtains

$$\frac{\partial^2 P_e}{\partial t^2} = \left[ P_2 \frac{1}{\sigma_2^2} \left( \frac{t - m_2}{\sigma_2} \right) \phi \left( \frac{t - m_2}{\sigma_2} \right) - P_1 \frac{1}{\sigma_1^2} \left( \frac{t - m_1}{\sigma_1} \right) \phi \left( \frac{t - m_1}{\sigma_1} \right) \right] \quad (42)$$

The root of equation (40) that gives a positive value for equation (42) is taken as the value of  $t$ , which minimizes the probability of error. Using the results of the last section, one can update the threshold  $t$  for use with the leave-one-out method since it is a function of means and covariance matrices.

##### 5. GENERALIZATION OF THE FISHER CLASSIFIER TO MULTIPLE CLASSES

Rewriting equations (12) and (13) in terms of the discriminant functions  $g_i(X) = V_i^T X + v_i$ ,  $i = 1, 2$ , the following decision rule is implemented:

$$\text{Decide } X \in \omega_1 \text{ if } g_1(X) > g_2(X) \quad (43)$$

$$\text{Decide } X \in \omega_2 \text{ if } g_1(X) < g_2(X) \quad (44)$$

Thus

$$V_i = \hat{S}_W^{-1} \hat{m}_i \quad (45)$$

and

$$v_i = -\hat{m}_i^T \hat{S}_W^{-1} \frac{(\hat{m}_1 + \hat{m}_2)}{2} \quad (46)$$

It is seen that equations (43) to (46) implement the decision rule of equation (6). This suggests the definition of discriminant functions for an  $M$ -class problem as

$$g_i(X) = V_i^T X + v_i, \quad i = 1, 2, \dots, M$$

where

$$\left. \begin{aligned} V_i &= \hat{S}_W^{-1} \hat{m}_i \\ v_i &= -\hat{m}_i^T \hat{S}_W^{-1} \frac{(\hat{m}_1 + \hat{m}_2 + \dots + \hat{m}_M)}{M}, \quad i = 1, 2, \dots, M \\ \hat{S}_W &= \hat{\Sigma}_1 + \hat{\Sigma}_2 + \dots + \hat{\Sigma}_M \end{aligned} \right\} \quad (47)$$

Then the decision rule is the following: Decide  $X \in \omega_i$  if

$$\begin{aligned} g_i(X) &> g_j(X) \\ j &= 1, 2, \dots, M \\ &\neq i \end{aligned} \quad (48)$$

## 6. COMPUTATIONAL EXPRESSIONS FOR THE LEAVE-ONE-OUT METHOD IN A MULTICLASS CASE

This section presents computational expressions for the leave-one-out method for updating  $V_i$  and  $v_i$ . Let there be  $M$  classes. Consider the case when a pattern  $x_k^1$  from class  $\omega_1$  is left out. Define the means and covariance matrices of the total pattern set as

$$\left. \begin{aligned} \hat{m}_i &= \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^i, \quad i = 1, 2, \dots, M \\ \hat{\Sigma}_1 &= \frac{1}{(N_1 - 2)} \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_1)(x_j^1 - \hat{m}_1)^T \\ \hat{\Sigma}_i &= \frac{1}{(N_i - 1)} \sum_{j=1}^{N_i} (x_j^i - \hat{m}_i)(x_j^i - \hat{m}_i)^T, \quad i = 2, \dots, M \end{aligned} \right\} \quad (49)$$

Let  $\hat{S}_W = \hat{\Sigma}_1 + \hat{\Sigma}_2 + \dots + \hat{\Sigma}_M$ . Compute  $V_i$  and  $v_i$ ,  $i = 1, 2, \dots, M$  as

$$\left. \begin{aligned} V_i &= \hat{S}_W^{-1} \hat{m}_i \\ v_i &= -\hat{m}_i^T \hat{S}_W^{-1} \frac{(\hat{m}_1 + \hat{m}_2 + \dots + \hat{m}_M)}{M} \end{aligned} \right\} \quad (50)$$

When the pattern  $x_k^1$  from class  $\omega_1$  is left out, Fisher's parameters are computed as

$$\left. \begin{aligned} V_1(x_k^1) &= \hat{S}_{W1k}^{-1} \hat{m}_{1k} \\ v_1(x_k^1) &= -\hat{m}_{1k}^T \hat{S}_{W1k}^{-1} \frac{(\hat{m}_{1k} + \hat{m}_2 + \dots + \hat{m}_M)}{M} \\ V_i(x_k^1) &= \hat{S}_{W1k}^{-1} \hat{m}_i, \quad i = 2, \dots, M \\ v_i(x_k^1) &= -\hat{m}_i^T \hat{S}_{W1k}^{-1} \frac{(\hat{m}_{1k} + \hat{m}_2 + \dots + \hat{m}_M)}{M}, \quad i = 2, \dots, M \end{aligned} \right\} \quad (51)$$

where

$$\left. \begin{aligned} \hat{m}_{1k} &= \frac{1}{(N_1 - 1)} \sum_{\substack{j=1 \\ \neq k}}^{N_1} x_j^1 \\ \hat{\Sigma}_{1k} &= \frac{1}{(N_1 - 2)} \sum_{\substack{j=1 \\ \neq k}}^{N_1} (x_j^1 - \hat{m}_1)(x_j^1 - \hat{m}_1)^T \\ \hat{S}_{W1k} &= \hat{\Sigma}_{1k} + \hat{\Sigma}_2 + \dots + \hat{\Sigma}_M \end{aligned} \right\} \quad (52)$$



The  $\hat{m}_i$  and  $\hat{\Sigma}_i$ ,  $i = 2, \dots, M$  are defined as in equation (49). Proceeding as in section 3.1, one obtains recursive relations for Fisher's parameters as follows:

$$\left. \begin{aligned} v_1(x_k^1) &= v_1 + \frac{\alpha d_1}{v(x_k^1)} y(x_k^1) - \frac{1}{(N_1 - 1)v(x_k^1)} y(x_k^1) \\ v_1(x_k^1) &= v_1 - \frac{\alpha}{v(x_k^1)} d_1 d_2 + \frac{d_2}{(N_1 - 1)v(x_k^1)} + \frac{d_1}{2(N_1 - 1)} \\ &\quad + \frac{\alpha \beta(x_k^1)}{2(N_1 - 1)v(x_k^1)} d_1 - \frac{1}{2(N_1 - 1)^2} \frac{\beta(x_k^1)}{v(x_k^1)} \end{aligned} \right\} \quad (53)$$

$$v_i(x_k^1) = v_i + \frac{\alpha e_i}{v(x_k^1)} y(x_k^1), \quad i = 2, \dots, M$$

(54)

$$v_i(x_k^1) = v_i - \frac{\alpha e_i d_2}{v(x_k^1)} + \frac{e_i}{2(N_1 - 1)v(x_k^1)}, \quad i = 2, \dots, M$$

where

$$\begin{aligned}
 \alpha &= \frac{N_1}{(N_1 - 1)(N_1 - 2)} \\
 \gamma(X_k^1) &= \hat{S}_W^{-1}(X_k^1 - \hat{m}_1) \\
 \beta(X_k^1) &= (X_k^1 - \hat{m}_1)^T \hat{S}_W^{-1}(X_k^1 - \hat{m}_1) \\
 v(X_k^1) &= 1 - \alpha\beta(X_k^1) \\
 d_1 &= \gamma^T(X_k^1) \hat{m}_1 \\
 d_2 &= \gamma^T(X_k^1) \frac{(\hat{m}_1 + \hat{m}_2 + \dots + \hat{m}_M)}{M} \\
 &= \gamma^T(X_k^1) \hat{m} \\
 e_i &= \gamma^T(X_k^1) \hat{m}_i, \quad i = 2, \dots, M
 \end{aligned} \tag{55}$$

Recursive relations can be obtained similarly when a pattern  $X_k^i$  from class  $\omega_i$  is left out. It is to be noted that the matrix  $\hat{S}_W$  is to be inverted once for each class. The use of these recursive relations results in a computationally efficient way of implementing the leave-one-out method.

## 7. CONCLUSIONS

The Fisher classifier is one of the simplest and most widely used linear classifiers. Recently, considerable interest in its application for the classification of multispectral data acquired by Landsat has been expressed. Acquiring labels of the training patterns is expensive, and in many cases the probability of error is to be estimated in addition to the designing of a classifier. (For example in remote sensing, a separate set of labeled patterns is used for estimating the probability of error.) Hence, in practical applications, it is advantageous to use the available labeled patterns more effectively.

This paper has presented computational expressions for estimating the probability of error using the leave-one-out method. Thus, the available labeled patterns can be used effectively, both for designing the classifier and estimating the probability of error. Since the classification accuracy depends on the threshold used with the Fisher classifier, expressions for optimal threshold for minimizing the probability of error in Fisher's direction are presented.

APPENDIX A  
DERIVATION OF MATRIX RELATIONS

From equation (14), one obtains

$$\begin{aligned}
 \hat{m}_{1k} &= \frac{1}{(N_1 - 1)} \left( \sum_{\substack{j=1 \\ \neq k}}^{N_1} x_j^1 \right) \\
 &= \frac{1}{(N_1 - 1)} \left( \sum_{j=1}^{N_1} x_j^1 - x_k^1 \right) \\
 &= \frac{N_1}{N_1 - 1} \hat{m}_1 - x_k^1 \\
 &= \hat{m}_1 - \frac{1}{(N_1 - 1)} (x_k^1 - \hat{m}_1)
 \end{aligned} \tag{A-1}$$

thus obtaining equation (18). From equation (15),

$$\begin{aligned}
 \hat{\Sigma}_{1k} &= \frac{1}{N_1 - 2} \sum_{\substack{j=1 \\ \neq k}}^{N_1} (x_j^1 - \hat{m}_{1k})(x_j^1 - \hat{m}_{1k})^T \\
 &= \frac{1}{N_1 - 2} \left[ \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_{1k})(x_j^1 - \hat{m}_{1k})^T - (x_k^1 - \hat{m}_{1k})(x_k^1 - \hat{m}_{1k})^T \right]
 \end{aligned} \tag{A-2}$$

Consider the following:

$$\begin{aligned}
\sum_{j=1}^{N_1} (x_j^1 - \hat{m}_{1k}) (x_j^1 - \hat{m}_{1k})^T &= \sum_{j=1}^{N_1} \left[ x_j^1 - \hat{m}_1 + \frac{1}{N_1 - 1} (x_k^1 - \hat{m}_1) \right] \left[ x_j^1 \right. \\
&\quad \left. - \hat{m}_1 + \frac{1}{N_1 - 1} (x_k^1 - \hat{m}_1) \right]^T \\
&= \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_1) (x_j^1 - \hat{m}_1)^T \\
&\quad + \frac{1}{(N_1 - 1)} (x_k^1 - \hat{m}_1) \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_1)^T \\
&\quad + \left[ \sum_{j=1}^{N_1} (x_j^1 - \hat{m}_1) \right] \frac{1}{(N_1 - 1)} (x_k^1 - \hat{m}_1)^T \\
&\quad + \frac{N_1}{(N_1 - 1)^2} (x_k^1 - \hat{m}_1) (x_k^1 - \hat{m}_1)^T \\
&= (N_1 - 2) \hat{\Sigma}_1 + \frac{N_1}{(N_1 - 1)^2} (x_k^1 - \hat{m}_1) (x_k^1 - \hat{m}_1)^T \quad (A-3)
\end{aligned}$$

Consider

$$\begin{aligned}
(x_k^1 - \hat{m}_{1k}) &= x_k^1 - \hat{m}_1 + \frac{1}{(N_1 - 1)} (x_k^1 - \hat{m}_1) \\
&= \frac{N_1}{N_1 - 1} (x_k^1 - \hat{m}_1) \quad (A-4)
\end{aligned}$$

Substituting equations (A-3) and (A-4) into (A-2) results in the following:

$$\begin{aligned}\hat{\Sigma}_{1k} &= \frac{1}{(N_1 - 2)} \left[ (N_1 - 2) \hat{\Sigma}_1 + \frac{N_1}{(N_1 - 1)^2} (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \right. \\ &\quad \left. - \frac{N_1^2}{(N_1 - 1)^2} (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \right] \\ &= \hat{\Sigma}_1 - \frac{N_1}{(N_1 - 1)(N_1 - 2)} \left[ (x_k^1 - \hat{m}_1)(x_k^1 - \hat{m}_1)^T \right]\end{aligned}\quad (A-5)$$

thus obtaining (19).

Let  $S = \Sigma - \alpha MM^T$ , where  $S$  and  $\Sigma$  are nonsingular matrices and  $M$  is a vector. Then the inverse of  $S$  can be expressed in terms of the inverse of  $\Sigma$  as in reference 6:

$$S^{-1} = \Sigma^{-1} + \frac{\alpha \Sigma^{-1} M M^T \Sigma^{-1}}{1 - \alpha M^T \Sigma^{-1} M}\quad (A-6)$$

thus obtaining equation (22).

## APPENDIX B

### REFERENCES

1. Lachenbruch, P. A.; and Mickey, M. R.: Estimation of Error Rates in Discriminant Analysis. *Technometrics*, vol. 10, Feb. 1968, pp. 1-11.
2. Fukunaga, K.; and Kessell, D. L.: Estimation of Classification Errors. *IEEE Trans. Computers*, vol. C-20, 1971, pp. 1521-1527.
3. Chittineni, C. B.: On the Estimation of Probability of Error. *Pattern Recognition*, vol. 9, no. 4, 1977, pp. 191-196.
4. Fisher, R. A.: The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugenics*, vol. 7, part II, 1936, pp. 179-188. [Also in contributions to *Mathematical Statistics*, John Wiley and Sons, Inc., New York, 1963.]
5. Chittineni, C. B.: On the Fisher Criterion and Divergence in Pattern Recognition. *Internat. J. Control*, vol. 16, no. 1, July 1972, pp. 205-207.
6. Bodewig, E.: *Matrix Calculus*. Amsterdam: North Holland, 1959.
7. Hills, M.: Allocation Rules and Their Error Rates. *J. Royal Stat. Soc., Ser. B.*, vol. 28, Jan. 1966, pp. 1-31.